Journal of Science and Technology in Civil Engineering, HUCE (NUCE), 2023, 17 (1): 125–136

AN ANALYSIS OF THE RELATIVE VARIABLE IMPORTANCE TO FLOOD FATALITY USING A MACHINE LEARNING APPROACH

Luu Thi Dieu Chinh[a], Ha Thi Hang[b,*]

[a] *Faculty of Hydraulic Engineering, Hanoi University of Civil Engineering, 55 Giai Phong road, Hai Ba Trung district, Hanoi, Vietnam*
[b] *Faculty of Bridges and Road, Hanoi University of Civil Engineering, 55 Giai Phong road, Hai Ba Trung district, Hanoi, Vietnam*

Article history:
Received 19/12/2022, Revised 16/02/2023, Accepted 08/3/2023

Abstract

Vietnam is regularly and severely affected by flood events and there were nearly 14,000 dead people in 200 separate floods from 1989 to 2015. However, there have been limited studies specifically on flood-related mortality in Vietnam. This paper presents a longitudinal investigation of flood fatalities in Vietnam. More specifically, we use the available national disaster database and machine learning techniques to investigate the importance of different attributes of flood damage to the attribute of flood fatalities. The results show that the housing damage attribute significantly influences the fatality attribute, of which the weights are 0.45, 0.62, and 0.36 for the random forest, boosting, and multiple linear regression techniques, respectively. Thus, it is recommended that the proper policy prioritize housing improvements, establish evacuation plans, and develop a strategy for temporary flood shelters in flood-prone areas. Understanding how various components of flood damage are more likely to lead to fatalities analyzed in this study is critical for developing risk reduction strategies.

Keywords: flood fatalities; national disaster database; machine learning; variable importance; disaster risk reduction; Vietnam.

https://doi.org/10.31814/stce.nuce2023-17(1)-10

1. Introduction

According to the Global Disaster Database, Asia is the region most at risk from natural disasters like floods and storms since 3,620 such events occurred there between 1900 and 2016 and caused at least 8,085,516 deaths (Fig. 1). Natural hazards almost often have the greatest impact on developing economies [1–3]. The potential for human casualties is important to consider when assessing flood risk [4]. In contrast to the substantial research conducted on flood-related mortality in developed countries [5–9], very few studies have been conducted on the topic in developing countries [10], despite their higher fatality rates [11]. Several approaches were used to study flood-related mortality.

Penning-Rowsell et al. [12] developed a model to forecast flood-related deaths and injuries using flood hazard and exposure aspects. Jonkman and Vrijling [13] presented a technique to estimate the death toll in flood occurrences by analyzing flood hazard characteristics, evaluating the exposed

*Corresponding author. *E-mail address:* hanght@huce.edu.vn (Hang, H. T.)

population, and assessing the fatality rate. Many other studies examined the context and root causes of flood-related mortality [14–16]. These studies were conducted for particular places or events and needed information such as the reason for death, evacuation condition, gender, age, time of day, and coping ability. Studying the causes of flood-related deaths is also possible via the use of prediction models [17–19]. It is hard to apply these models in data-poor regions because of the exact input parameters they need when paired with hydraulic models. A lack of precise data for analysis contributes to the paucity of research on mortality caused by floods in developing nations.

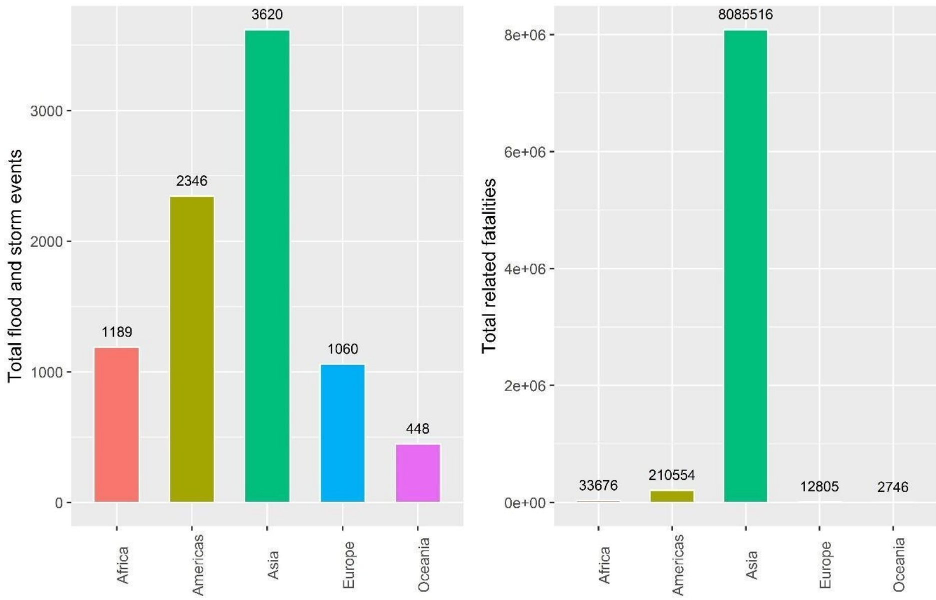


Figure 1. Worldwide flood and storm occurrences and fatalities from 1900 to 2016 (information retrieved on April 20, 2017, from http://emdat.be)

Analyzing disaster damage data may give important information for policy development and decision-making in disaster management [20, 21]. Disaster databases that are available for public access are grouped and aggregated in the publications of Simpson et al. [22] and Grasso and Dilley [23], such as the global disaster database (www.emdat.be) and national disaster databases (www.desinventar.net). These databases can be used for flood studies in which the data are analyzed and converted into useful information for disaster management. In addition, we also need to focus on research on data recording standards to improve the quality of disaster databases and to provide diverse and accurate databases for analysis [23].

Vietnam is significantly influenced by flood hazards, being the ninth most severely affected by extreme weather events [24] and the fourth most vulnerable to river flood danger by the percentage of the population [25]. Thus, there has been a significant growth in research on flood risk in Vietnam. Tran et al. [26] analyzed how rural Thua Thien Hue province people deal with the threat of flooding. Chau et al. [27] analyzed the impacts of flood hazards on the agriculture sector in Quang Nam province using GIS techniques. Chinh et al. [28] had a survey with households and small private businesses in Can Tho city on flood preparedness, response, and recovery. The machine learning method was utilized to model the vulnerability of the Tuong Duong district [29]. Luu et al. [30] used spatial analysis techniques to evaluate the risk of flooding in Quang Nam province. On the other hand, there have been no studies conducted thus far on the deaths caused by floods in Vietnam.

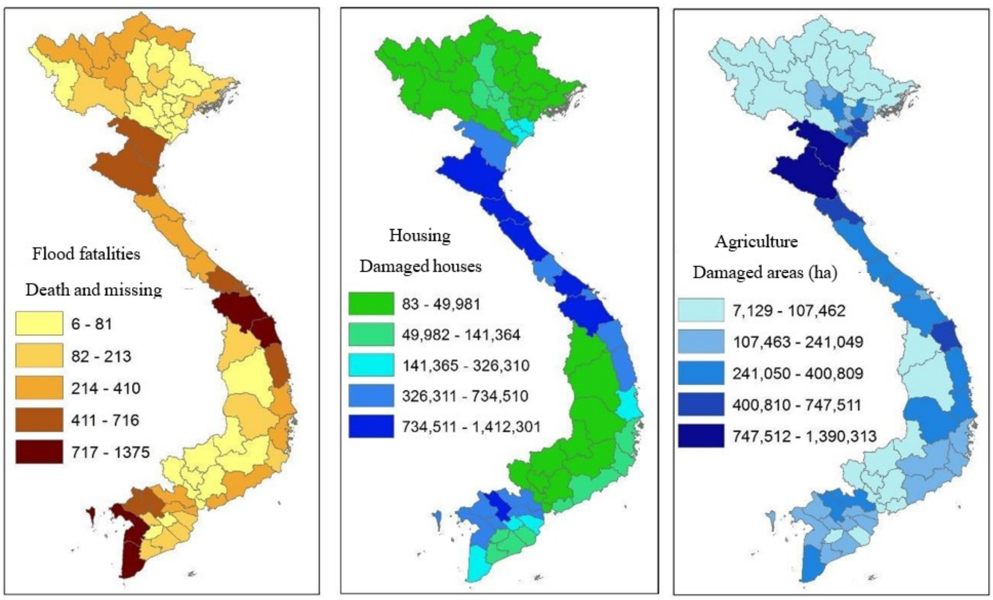
DANA (Damage Assessment and Needs Analysis) is the acronym for the Vietnam national disaster database. It was developed by the government commissioner of the Department of Dyke Management, Flood and Storm Control under the support of the United Nations Development Programme [23]. Since 1989, damages from disaster events have been recorded [31, 32]. Machine learning algorithmic models have been more and more popular in natural hazard assessment, including floods [33]. Machine learning techniques are very useful to solve complicated problems with multiple interacting factors and can analyze the importance of attended variables in flood studies [34, 35]. Besides, ML techniques can be arranged according to a diverse classification to denote the desired results of the modeling procedure [36]. Thus, this study aimed to analyze the DANA data (documented from 1989-2015) to assess the influencing attributes on fatality attributes. We used advanced machine learning techniques of random forest, multiple linear regression, and boosting. The result may help define flood risk management strategies and relevant measures to reduce mortality from future floods.

2. Material and Methods

2.1. Data used

Data on flood damage had been compiled by the National Steering Committee for Flood and Storm Control via the DANA system [31, 37, 38]. Like many other international, regional, and national disaster databases, DANA does not include data on the physical exposure of buildings and infrastructure [22]. DANA compiles information about disaster-related losses by classifying them according to many different attributes, such as the number of people killed or injured, the number of houses destroyed, the area of rice and crop lost, the amount of soil washed away, and the aquaculture areas [32]. However, the DANA database only records the values lost due to business interruption, not the indirect losses resulting from the recovery and restoration of destroyed properties and infrastructure [39].

We collected more than 200 flood and storm data cards between 1989 and 2015 from the DANA database. After that, the data are compiled for 63 provinces of Vietnam, as shown in Fig. 2.



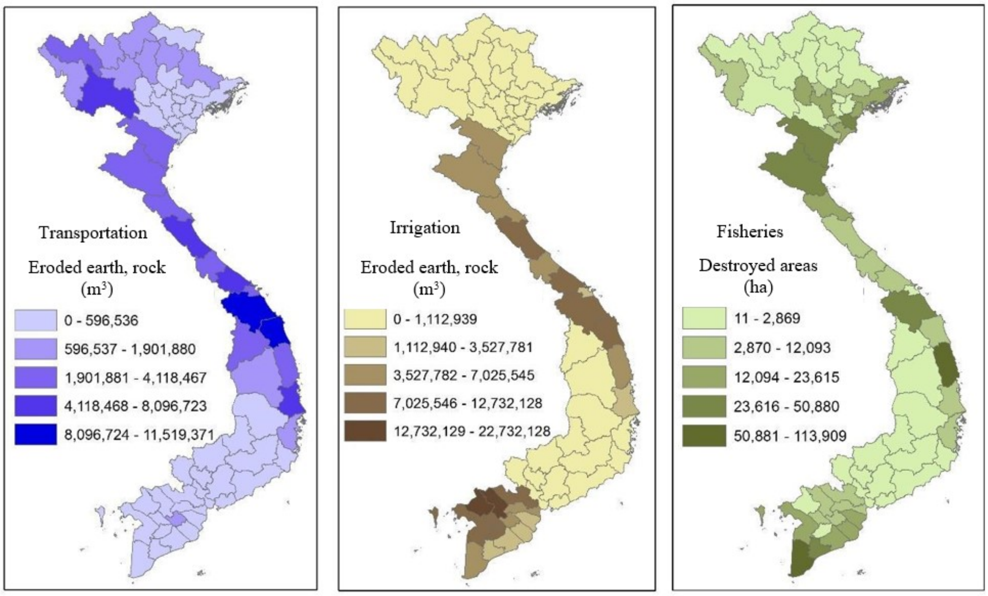


Figure 2. The spatial distribution of flood impacts at the provincial level

Flood fatality is a key indicator in risk analysis. The number of confirmed fatalities from a flood is known as the death toll [31]. The Vietnamese people have suffered greatly as a result of flood and storm disasters, particularly in terms of mortality. Between 1989 and 2015, there were at least 14,927 flood fatalities documented (Fig. 3).

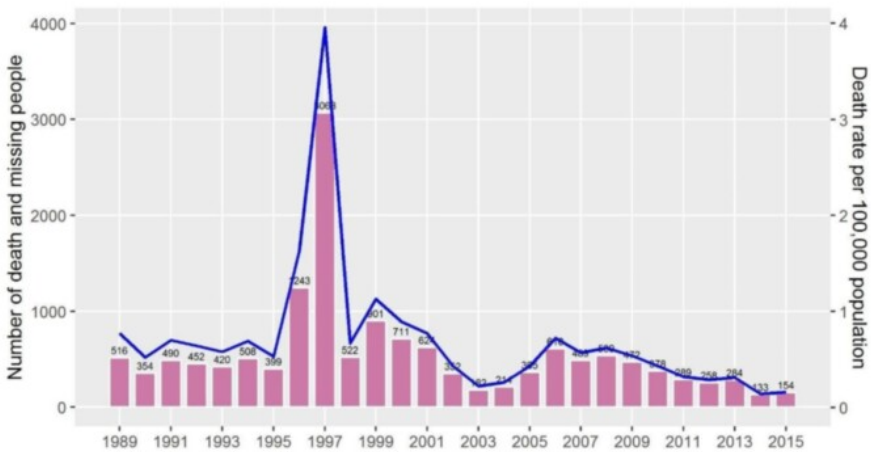


Figure 3. Temporal distribution of flood-related fatalities and death rate per 100,000 population in Vietnam, 1989-2015

The following flood damage attributes are considered for analysis: death toll, house damage, agriculture loss, transportation damage, irrigation damage, and fisheries loss (as shown in Fig. 2). The six variables of the flood damage attributes are collected for 63 provinces from 1989 to 2015. There are a total of 1,701 observations, one for every year for each province.

2.2. *Method used*

This study aims to examine the relationship between different flood damage qualities and the mortality attribute using the DANA database. We focus on measuring the variable importance using statistical machine learning models. Statistical tools in R software are used to analyze the relationship between flood fatality attribute and other flood damage attributes [40]. Fig. 4 details the analytic approach used in this study. There are five steps to this process: (1) collecting a sample data set, (2) splitting randomly the sample data set into training data set and testing data set, (3) constructing a training model, (4) applying it to a testing model for model validation, and (5) measuring model performance via an index, Mean Squared Error (MSE).

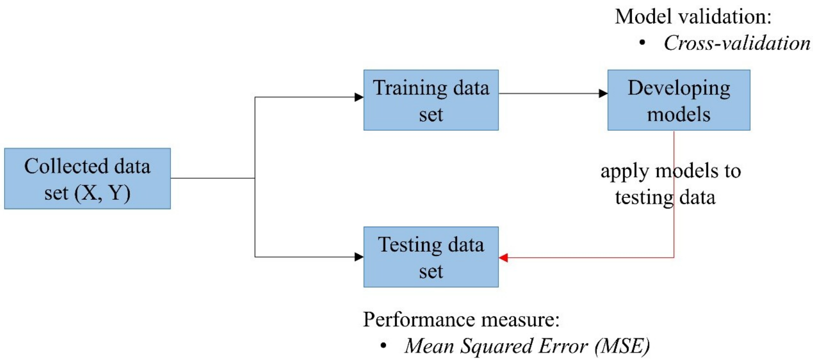


Figure 4. The machine learning framework applied in this study

We calculated the mean, median, range, minimum, and maximum for examining the statistical distributions of variables. Similar to the characteristic of data in the study of Zhou et al. [41], our flood damage data have a large variation. Therefore we applied log transformation for all six flood damage categories to better fit the normal distribution. We applied the machine learning framework in Fig. 4. The sample data is randomly split into training and testing data sets at a ratio of 70% and 30%, respectively, using the 'caTools' package in R software [42]. To verify the model, we used the cross-validation technique. In this procedure, the training data set was used to create the model and the testing data set to verify it.

We used three advanced machine learning models of random forest, boosting, and multiple linear regression to investigate the relationship between flood damage attributes and fatality attributes. These methods provide a simple and natural justification for how each variable affects the whole. The following sections elaborate on each of these models.

a. Multiple linear regression

A typical statistical modeling method known as multiple linear regression may be found in any modeling book, for example, the book [43]. The research uses multiple linear regression to model flood mortality as a linear function of other damage attributes. Assumed is the absence of any variable interactions. After that, we used LMG measure (the abbreviation of Lindeman, Merend, and Gold [44]) to compute the variable importance in the linear model. The LMG measure is based on the proportional contribution of the total R-squared of each variable in the linear model [44, 45].

b. Random forest

Random forest for regression is another method that is used on the same data. Random forest is a method for improving prediction accuracy by combining numerous separate weak regression models, such as decision trees, which often have high bias and low variance, into a single ensemble model [46]. This study establishes the random forest model using the 'randomForest' package in R software [47]. The random forest algorithm is based on Breiman et al. [46] 's classification and regression tree or CART.

c. Boosting

The boosting method, like the random forest method, combines many separate, weak regression models to create a more robust and precise ensemble model. Contrary to the random forest, boosting is a sequential strategy in which each weak model is introduced to improve the performance of the prior group of weak models. This research used the R 'gbm' package to construct a boosting model from decision trees [47].

3. Results

3.1. Variable importance measures using multiple linear regression

We conducted an analysis of the correlation between the log-transformed flood mortality attribute and other flood damage variables [40]. The machine learning framework in Fig. 4 was applied for data analysis. The residual standard error of the training model is 0.856, and its multiple coefficient of determination is 0.5932. The MSE of the model under testing is 0.674. The 'relaimpo' package is then used to create relative variable importance based on the LMG indicator [48].

The linear regression model was used to examine the correlation between the flood damage attributes and the flood death attribute. Training data sets are used to develop the model, while test data sets are used to test model performance. The model of the training data set has a defined coefficient (squared R) of 0.5932, and the remaining standard error is 0.856. Next, we calculated the forecasted value of the test data set model. The average square error (MSE) of the test model is 0.674. We then ran the 'relaimpo' package in R software to create significant change measures based on the LMG indicator [48].

According to the findings shown in Table 1, the feature representing housing damage is the most significant factor to consider since it has a relative importance of 0.3642. The significance of the other factors is not as great. The P value for each variable comes in at less than 0.05.

Table 1. Variable importance from multiple linear regression modeling

Damage attributes	Relative variable importance	P-value
Housing	0.3642	< 0.0001
Agriculture	0.1791	0.0468
Transportation	0.1548	< 0.0001
Irrigation	0.1527	< 0.0001
Fisheries	0.1492	0.0221

3.2. Variable importance measures using random forest

The influence of flood damage variables on the mortality attribute is quantified using a random forest regression model. The random forest model is constructed using the 'randomForest' package in R software [49]. The input parameters are set for the modeling, with 500 for the number of trees and

15 for the maximum number of terminal nodes. The mean of squared residuals for the training model is 0.687, and the percentage variance explained is 61.64%. The MSE of the testing model is 0.612.

Table 2 displays the random forest model's variable importance measures. The variable importance results of the random forest modeling show that housing damage is the most important variable, while other factors have a limited influence, which is consistent with the finding from the multiple linear regression modeling.

Table 2. Variable importance from random forest modeling

Damage attributes	Relative variable importance
Housing	0.4546
Irrigation	0.2438
Agriculture	0.1260
Fisheries	0.0906
Transportation	0.0848

3.3. *Variable importance measures using boosting*

We fitted the boosted regression model using the training data set and the 'gbm' package in R software [47]. The input parameters are set for the modeling with 10,000 for the number of trees, 4 for the maximum depth of variable interactions, and Gaussian for the distribution assumption. The performance of the boosting model is evaluated using the cross-validation method. Its MSE is 0.638 for the test data set. It is compatible with the results of the multiple linear regression model (0.674) and the random forest model (0.612). Small variations in MSEs among methods suggest the used models are accurate and well-validated.

Table 3 displays the outcomes of the boosting model's variable importance measurements. We also discover that the mortality attribute is most heavily influenced by the flood housing damage attribute and is only somewhat influenced by the other damage characteristics. The three used models all arrive at similar conclusions throughout the analysis.

Table 3. Variable importance from boosting modeling

Damage attributes	Relative variable importance
Housing	0.6191
Irrigation	0.1372
Agriculture	0.0959
Fisheries	0.0845
Transportation	0.0633

4. Discussion

Table 4 summarizes the findings of variable significance measurements for three applicable models. There is consensus between the models that the house damage attribute has the greatest impact on the flood fatality attribute, whereas other damage qualities have considerably less impact. The outcomes of the boosting and random forest models are comparable for all variables. The linear regression model and several other models have somewhat different rankings for some aspects of damage. The variations may be traced back to the underlying assumptions of the statistical methods used. In contrast to the random forest and boosting models, which are done under the premise of interaction between variables, the multiple linear regression model assumes that there are no such interactions [43].

Table 4. Comparison of variable importance measures between used models

Damage attribute/ Statistical model	Relative importance measures					
	Multiple linear regression		Random forest		Boosting	
	Weight	Rank	Weight	Rank	Weight	Rank
Housing	0.3642	1	0.4546	1	0.6191	1
Irrigation	0.1527	4	0.2438	2	0.1372	2
Agriculture	0.1791	2	0.1260	3	0.0959	3
Fisheries	0.1492	5	0.0906	4	0.0845	4
Transportation	0.1548	3	0.0848	5	0.0633	5

Our findings show the high significance of the flood housing damage attribute in relation to the fatality attribute. This result is supported by the observation of De Bruijn and Klijn [50]. Collapsing buildings were noted as one of the major causes of mortality during floods by De Bruijn and Klijn [50]. However, a significant number of deaths involving vehicles have been associated with flood occurrences in several developed countries [6, 7, 9]. In Europe and the US, automobiles were determined to be responsible for 38.5% of fatalities, whereas buildings only accounted for 9.3% [13]. 63% of flood-related deaths in the United States were found to be vehicle-related [6, 9]. 48.5% of flood-related deaths in Australia included vehicles [7]. As a result, economic, social, and income situations are probably connected to variables determining flood fatalities. The quality of housing stock in developing countries is a massive risk to life. The current research could provide a more robust basis for Vietnam's flood risk reduction decision-making procedures.

The results are a true reflection of Vietnamese dwelling conditions in flood-prone locations. As shown in some representative instances in Fig. 5, the majority of the dwellings in the rural region are single-story and in bad condition. They are often severely damaged or destroyed by storms or heavy rain events. Flood depths of more than two meters are common in many floodplain areas in central Vietnam [30]. Residents in floodplains susceptible to high flood depth levels should not live in single-story homes. Several appropriate measures can be applied to these situations, such as amphibious houses [51] and evacuation plans [52].

The most vulnerable populations in rural Vietnam are those who are impoverished or marginalized. They struggle greatly because they have limited access to public resources like insurance and emergency assistance [53]. Under Decree 67/2007/ND-CP, the affected families sometimes get financial assistance from the government; nonetheless, this financing is very little compared to the demand. As is customary worldwide, the effect of disasters is inextricably related to the poverty rate. Addressing societal vulnerability, and therefore flood risk and mortality tolls requires more focus.

Consideration must be given to potential solutions to reduce flood-related fatalities. These actions should reduce flood risk throughout the long period of growth and in the near term. Three policy implications are suggested based on the current study findings to provide decision-makers with something to consider when prioritizing flood risk control operations and appropriately allocating resources.

(a) A house collapsed in 2006 Typhoon Xangsane [54]

(b) Houses in Central Vietnam were inundated in November of 2016 [55]

Figure 5. Some typical dwellings in flood-prone areas in Vietnam

First, it is crucial to improve the house quality in flood-prone regions of Vietnam, as shown by the strong influence of the flood dwelling damage attribute on the mortality attribute. Therefore, assistance for the poor to modify their homes for flood risk adaptation should be given top priority in government programs on disaster risk reduction. The government should work with the locals on the rehabilitation project and take solutions such as amphibious houses [51] and 2-floor resistant dwellings [54].

Second, flood deaths might be considerably decreased with proper evacuation arrangements [56]. To reduce damage from catastrophes, communities' capacity to evacuate during storm and flood events has to be strengthened [52, 57]. It is vital to create evacuation plans for dangerous regions based on integrating flood hazard assessments with data on housing and population since the housing factor substantially impacts flood fatalities in Vietnam.

Third, in low-income nations with subpar engineering building performance and maintenance, flood shelter is one of the most effective methods to prevent flood-related mortality after evacuation preparations [52]. Under the funding of NGOs, flood shelters have been built in a number of flood-prone localities in Vietnam [58, 59]. Government strategies on disaster risk mitigation should prioritize building flood shelters in flood-affected regions due to the important impact of flood home damage on fatality and the efficacy of flood shelters.

Since the advent of big data, machine learning methods have been widely used in academia and industry. This is the first time that machine learning techniques and a national catastrophe database have been used to conduct longitudinal research on the correlation between flood damage and fatalities. Our analysis provides a more detailed picture of flood fatalities. It is possible that this research, which uses a machine learning approach, may strengthen the machine-learning approach for flood risk management and encourage the use of disaster databases for policy applications.

5. Conclusions

Based on data from Vietnam's national disaster database, this report offers an informative analysis of flood mortality in Vietnam from 1989 to 2015. Relative predictor relevance is calculated using statistical machine learning methods. The flood housing damage attribute has the greatest relevance on the flood mortality attribute in Vietnam, according to the findings of variable importance measurements. Our results have applications for governmental initiatives that aim to reduce human casualties.

Some policy suggestions include raising the standard of poor people's homes, creating evacuation plans for places with a high danger of flooding, and building flood shelters in flood-prone areas. Using national and regional catastrophe datasets, where damages are classified by characteristics and recorded for a long time, machine learning techniques may be used to investigate the variable importance measures in different nations and areas.

This research is restricted to looking at how flood damage characteristics compare to mortality attributes in the national disaster database of Vietnam. This restriction relates to the recorded national disaster database, which is not equipped with the data to represent physical damage. To decrease fatalities in future flood occurrences, government strategies in flood risk management may benefit from knowing the relative impact measurements of flood damage attributes to fatality attributes. Future studies should make an effort to include additional predictor factors, such as flood hazard type, exposure, and demographic data (victim age and gender), to better understand the causes of flood deaths in Vietnam.

6. Acknowledgments

We sincerely thank the Central Steering Committee for Natural Disaster Prevention and Control in Vietnam for providing the flood damage data.

References

[1] Kahn, M. E. (2005). The Death Toll from Natural Disasters: The Role of Income, Geography, and Institutions. *Review of Economics and Statistics*, 87(2):271–284.

[2] Hansson, K., Danielson, M., Ekenberg, L. (2008). A framework for evaluation of flood management strategies. *Journal of Environmental Management*, 86(3):465–480.

[3] Jongman, B., Winsemius, H. C., Aerts, J. C. J. H., de Perez, E. C., van Aalst, M. K., Kron, W., Ward, P. J. (2015). Declining vulnerability to river floods and the global benefits of adaptation. *Proceedings of the National Academy of Sciences*, 112(18).

[4] Chinh, L. T. D., Hằng, H. T., Linh, T. H., Quỳnh, B. D. (2022). Assessing flood impacts for Quang Binh province using Google Earth Engine and spatial analyses. *Journal of Science and Technology in Civil Engineering (STCE) - HUCE*, 16(5V):101–113.

[5] Coates, L. (1999). Flood Fatalities in Australia, 1788-1996. *Australian Geographer*, 30(3):391–408.

[6] Ashley, S. T., Ashley, W. S. (2008). Flood Fatalities in the United States. *Journal of Applied Meteorology and Climatology*, 47(3):805–818.

[7] FitzGerald, G., Du, W., Jamal, A., Clark, M., Hou, X.-Y. (2010). Flood fatalities in contemporary Australia (1997-2008). *Emergency Medicine Australasia*, 22(2):180–186.

[8] Sharif, H. O., Jackson, T. L., Hossain, M. M., Zane, D. (2015). Analysis of Flood Fatalities in Texas. *Natural Hazards Review*, 16(1).

[9] Terti, G., Ruin, I., Gourley, J. J., Kirstetter, P., Flamig, Z., Blanchet, J., Arthur, A., Anquetin, S. (2017). Toward Probabilistic Prediction of Flash Flood Human Impacts. *Risk Analysis*, 39(1):140–161.

[10] Paul, B. K., Mahmood, S. (2016). Selected physical parameters as determinants of flood fatalities in Bangladesh, 1972–2013. *Natural Hazards*.

[11] Jonkman, S. N. (2005). Global Perspectives on Loss of Human Life Caused by Floods. *Natural Hazards*, 34(2):151–175.

[12] Penning-Rowsell, E., Floyd, P., Ramsbottom, D., Surendran, S. (2005). Estimating Injury and Loss of Life in Floods: A Deterministic Framework. *Natural Hazards*, 36(1-2):43–64.

[13] Jonkman, S., Vrijling, J. (2008). Loss of life due to floods. *Journal of Flood Risk Management*, 1(1): 43–56.

[14] Jonkman, S. N., Kelman, I. (2005). An Analysis of the Causes and Circumstances of Flood Disaster Deaths. *Disasters*, 29(1):75–97.

[15] Jonkman, S. N., Maaskant, B., Boyd, E., Levitan, M. L. (2009). Loss of Life Caused by the Flooding of New Orleans After Hurricane Katrina: Analysis of the Relationship Between Flood Characteristics and Mortality. *Risk Analysis*, 29(5):676–698.

[16] Haynes, K., Coates, L., van den Honert, R., Gissing, A., Bird, D., de Oliveira, F. D., D'Arcy, R., Smith, C., Radford, D. (2017). Exploring the circumstances surrounding flood fatalities in Australia—1900–2015 and the implications for policy and practice. *Environmental Science & Policy*, 76:165–176.

[17] Zhai, G., Fukuzono, T., Ikeda, S. (2006). An empirical model of fatalities and injuries due to floods in Japan. *Journal of the American Water Resources Association*, 42(4):863–875.

[18] Mauro, M. D., de Bruijn, K. M. (2012). Application and validation of mortality functions to assess the consequences of flooding to people. *Journal of Flood Risk Management*, 5(2):92–110.

[19] Mauro, M. D., Bruijn, K. M. D., Meloni, M. (2012). Quantitative methods for estimating flood fatalities: towards the introduction of loss-of-life estimation in the assessment of flood risk. *Natural Hazards*, 63 (2):1083–1113.

[20] UNISDR (2015). *Global Assessment Report on Disaster Risk Reduction 2015*.

[21] IRDR (2014). *IRDR Peril Classification and Hazard Glossary*. Integrated Research on Disaster Risk (IRDR), Beijing.

[22] Simpson, A., Murnane, R., Saito, K., Phillips, E., Reid, R., Himmelfarb, A. (2014). *Understanding Risk in an Evolving World: Emerging Best Practices in Natural Disaster Risk Assessment. Washington DC*. Washington DC: Global Facility for Disaster Reduction and Recovery, the World Bank.

[23] Grasso, V. F., Dilley, M. (2013). *A Comparative Review of Country-Level and Regional Disaster Loss and Damage Databases*. United Nations Development Programme Bureau for Crisis Prevention and Recovery, New York.

[24] Kreft, S., Eckstein, D., Melchior, I. (2016). *Global climate risk index 2017*. Berlin: Germanwatch e.V.

[25] Luo, T., Maddocks, A., Iceland, C., Ward, P., Winsemius, H. (2015). *World's 15 countries with the most people exposed to river floods*.

[26] Tran, P., Marincioni, F., Shaw, R., Sarti, M., An, L. V. (2007). Flood risk management in Central Viet Nam: challenges and potentials. *Natural Hazards*, 46(1):119–138.

[27] Chau, V. N., Holland, J., Cassells, S., Tuohy, M. (2013). Using GIS to map impacts upon agriculture from extreme floods in Vietnam. *Applied Geography*, 41:65–74.

[28] Chinh, D. T., Bubeck, P., Dung, N. V., Kreibich, H. (2016). The 2011 flood event in the Mekong Delta: preparedness, response, damage and recovery of private households and small businesses. *Disasters*, 40 (4):753–778.

[29] Bui, D. T., Pradhan, B., Nampak, H., Bui, Q.-T., Tran, Q.-A., Nguyen, Q.-P. (2016). Hybrid artificial intelligence approach based on neural fuzzy inference model and metaheuristic optimization for flood susceptibilitgy modeling in a high-frequency tropical cyclone area using GIS. *Journal of Hydrology*, 540: 317–330.

[30] Luu, C., Meding, J. V., Kanjanabootra, S. (2017). Assessing flood hazard using flood marks and analytic hierarchy process approach: a case study for the 2013 flood event in Quang Nam, Vietnam. *Natural Hazards*, 90(3):1031–1050.

[31] Below, R., Vos, F., Guha-Sapir, D. (2010). *Moving towards harmonization of disaster data: a study of six Asian databases*. Brussels: Centre for Research on the Epidemiology of Disasters.

[32] Hughey, E., Bell, H., Chatman, M. (2011). Who Needs What? A Case Study of Post-disaster Damage and Needs Assessment (DANA) in Vietnam. *Risk, Hazards & Crisis in Public Policy*, 2(4):1–24.

[33] Tehrany, M. S., Jones, S., Shabani, F. (2019). Identifying the essential flood conditioning factors for flood prone area mapping using machine learning techniques. *CATENA*, 175:174–192.

[34] Wang, Z., Lai, C., Chen, X., Yang, B., Zhao, S., Bai, X. (2015). Flood hazard risk assessment model based on random forest. *Journal of Hydrology*, 527:1130–1141.

[35] Zhao, G., Pang, B., Xu, Z., Yue, J., Tu, T. (2018). Mapping flood susceptibility in mountainous areas on a national scale in China. *Science of The Total Environment*, 615:1133–1142.

[36] Rahmati, O., Yousefi, S., Kalantari, Z., Uuemaa, E., Teimurian, T., Keesstra, S., Pham, T., Bui, D. T. (2019). Multi-Hazard Exposure Mapping Using Machine Learning Techniques: A Case Study from Iran. *Remote Sensing*, 11(16):1943.

[37] MARD (2006). *Guideline on Natural Disaster Damage and Needs Assessment*. UNDP Project VIE/01/014 - Capacity Building for Disaster Mitigation in Vietnam.

[38] Bollin, C., Khanna, S. (2007). *Review of Post Disaster Recovery Needs Assessment and Methodologies: Experiences from Asia and Latin America*. International Recovery Platform.

[39] Wang, X., Mahul, O., Stutley, C. (2010). *Weathering the Storm: Options for Disaster Risk Financing in Vietnam*. Washington, DC: Global Facility for Disaster Reduction and Recovery, the World Bank.

[40] R Core Team. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

[41] Zhou, Q., Leng, G., Feng, L. (2017). Predictability of state-level flood damage in the conterminous United States: the role of hazard, exposure and vulnerability. *Scientific Reports*, 7(1).

[42] Tuszynski, J. (2014). *caTools: Tools: moving window statistics, GIF, Base64, ROC AUC, etc.. R package version 1.17.1*.

[43] James, G., Witten, D., Hastie, T., Tibshirani, R. (2013). *An Introduction to Statistical Learning*. Springer New York.

[44] Lindeman, R. H., Merenda, P. F. (1980). *Introduction to bivariate and multivariate analysis*. The USA: Scott, Foresman and Company.

[45] Grömping, U. (2015). Variable importance in regression models. *Wiley Interdisciplinary Reviews: Computational Statistics*, 7(2):137–152.

[46] Breiman, L., Friedman, J., Stone, C. J., Olshen, R. A. (1984). *Classification and regression trees*. Monterey: Wadsworth.

[47] Ridgeway, G. (2015). *gbm: generalized boosted regression models. R package version 2.1.1*.

[48] Grömping, U. (2006). Relative Importance for Linear Regression in R: The Package relaimpo. *Journal of Statistical Software*, 17(1).

[49] Liaw, A., Wiener, M. (2002). Classification and regression by randomForest. *R news*, 2(3):18–22.

[50] de Bruijn, K. M., Klijn, F., van de Pas, B., Slager, C. T. J. (2015). Flood fatality hazard and flood damage hazard: combining multiple hazard characteristics into meaningful maps for spatial planning. *Natural Hazards and Earth System Sciences*, 15(6):1297–1309.

[51] English, E. C., Friedland, C. J., Orooji, F. (2017). Combined Flood and Wind Mitigation for Hurricane Damage Prevention: Case for Amphibious Construction. *Journal of Structural Engineering*, 143(6).

[52] Masuya, A., Dewan, A., Corner, R. J. (2015). Population evacuation: evaluating spatial distribution of flood shelters and vulnerable residential units in Dhaka with geographic information systems. *Natural Hazards*, 78(3):1859–1882.

[53] Chau, V. N., Holland, J., Cassells, S. (2014). Institutional structures underpinning flood management in Vietnam. *International Journal of Disaster Risk Reduction*, 10:341–348.

[54] Tran, T. A. (2016). *Developing Disaster Resilient Housing in Vietnam: Challenges and Solutions*. Springer International Publishing.

[55] Luu, C., Meding, J. V., Kanjanabootra, S. (2017). Analysing flood fatalities in Vietnam using national disaster database and tree-based methods.

[56] Coppola, D. P. (2015). *Introduction to International Disaster Management*. Boston: Butterworth-Heinemann.

[57] Liu, D., Li, Y., Fang, S., Zhang, Y. (2017). Influencing factors for emergency evacuation capability of rural households to flood hazards in western mountainous regions of Henan province, China. *International Journal of Disaster Risk Reduction*, 21:187–195.

[58] Anh, T. T., Phong, T. V. G., Mulenga, M. (2014). Community consultation for climate resilient housing: A comparative case study in Vietnam. *International Journal of Disaster Risk Reduction*, 10:201–212.

[59] Tran, T. A. (2015). Post-disaster housing reconstruction as a significant opportunity to building disaster resilience: a case in Vietnam. *Natural Hazards*, 79(1):61–79.